

Mario Fischer

»WDF*IDF?

Jetzt mal Butter bei die Fische!

WDF, IDF, Termgewichtung, Terminalsignale, Diversity und Holismus, Dokumentkorpus ... Hört sich alles wunderbar kompliziert und damit automatisch extrem wichtig an. Aber was macht man als Websitebetreiber denn jetzt damit genau? Geht es wirklich nicht ohne ein entsprechendes Tool? Wir versuchen, möglichst einfache Antworten auf komplizierte Fragen zu geben und zu zeigen, wie man Texte hinsichtlich ihrer Termgewichtung in der täglichen Praxis weiter optimieren kann. Ob der aktuelle Hype um die neue SEO-Formel gerechtfertigt erscheint, darüber sollte sich jeder selbst ein Bild machen. Wie immer gilt: Probieren geht über Studieren!

Ja, Text ist für Suchmaschinen extrem wichtig. Und ja: Natürlich ist es nicht egal oder zweitrangig, welche Worte wo auf einer Webseite stehen. Suchmaschinen, allen voran Google, verwenden mittlerweile nicht nur sehr komplexe Algorithmen zur Inhaltserkennung und -bestimmung, sondern ziehen auch den Vergleich mit anderen, ähnlichen Seiten zurate. Dieser technischen Textanalyse kann man in professioneller Weise auch eine technische Textoptimierung entgegensetzen.

Die Theorie

In der Theorie geht man davon aus, dass z. B. die ersten zehn Treffer zu einem Suchterm – also ein oder mehrere Worte – bei Google unter anderem dort stehen, weil ihre Texte bestimmte Merkmale aufweisen. Der eigentliche Suchterm sollte nach der WDF*IDF-Formel den höchsten Wert alle Terme haben und idealerweise sind auch weitere „beweisführende“ Terme in möglichst vielen dieser Dokumente vorhanden. Die Abbildung 1 zeigt dieses Prinzip stark vereinfacht auf: Angenommen, für das Suchwort „Westernstiefel“ findet man in den ersten fünf Treffern („nur“ fünf Treffer wurden für eine übersichtlichere Darstellung gewählt) bei Google ebenfalls

die Terme „Schuhe“, „Naht“ und „Leder-schaft“ in einer hohen Gewichtung, dann kann man davon ausgehen, dass sie in gewisser Weise zum Term „Westernstiefel“ gehören.

Das ist nichts anderes als die kreative Abwandlung eines alten Sprichwortes: „Wer A sagt, sagt häufig auch B.“ Wenn Menschen sich über Rasen unterhalten, fallen häufig auch die Worte „grün“ oder „Rasenmäher“. Fallen hingegen Worte wie „Blitzer“, „Punkte“ und „Flensburg“, hat der Begriffsgebrauch „Rasen“ eine völlig andere semantische Bedeutung. Suchmaschinen können über solche „Semantic Spaces“ Texte zum Auto Golf von der Sportart Golf und dem Golf von Mexiko unterscheiden. In je höherer Gewichtung diese „Beiworte“ neben dem eigentlichen Hauptterm vorkommen – und in je mehr passenden Dokumenten –, desto stärker ist ihre Beweiskraft. Viele Ärzte haben in Krankenhaus einen weißen Kittel an und tragen für alle gut sichtbar einen Pieper. Und natürlich diese weißen Plastik-Cloggs. Hängt jetzt noch ein Stethoskop um den Hals und auf dem Namensschild steht ein „Dr.“ vor den Namen, handelt es sich fast 100% Sicherheit um einen Arzt. Niemand käme hier auf die Idee, der Kittelträger könnte Dr. der Ma-

thematik sein. Genau so „denken“ im Prinzip die Algorithmen der Suchmaschine. Wäre der Kittel nicht weiß oder OP-grün, sondern z. B. orange, hätten wir sicher starke Zweifel, ob wir einen Arzt oder jemand von den Müllentsorgern vor uns hätten.

Die Praxis

Wie kleidet man sich also als Arzt im Krankenhaus, wenn man bei den Patienten Vertrauen erwecken will? Richtig! Weißer Kittel, Pieper, Cloggs, Namensschild und Stethoskop. Ohne eine #Sexismusdebatte auslösen zu wollen, würde dies auch eine blutjunge Assistenzärztin davor bewahren, mit einer Krankenschwester verwechselt zu werden, denn bei der fehlt Pieper, Stethoskop und bei wirklich genauerem Hinsehen auch das „Dr.“ (Randbemerkung: Wem man als Patient mehr vertrauen sollte, bleibt an dieser Stelle offen.) Wären beide ein typisches Dokument, könnte man sie damit recht gut auseinander halten. Wenn dem im realen Leben so ist, spricht dann nicht vieles dafür, ein echtes Dokument im Web textuell möglichst exakt so auszu-

„Wer A sagt, sagt häufig auch B!“

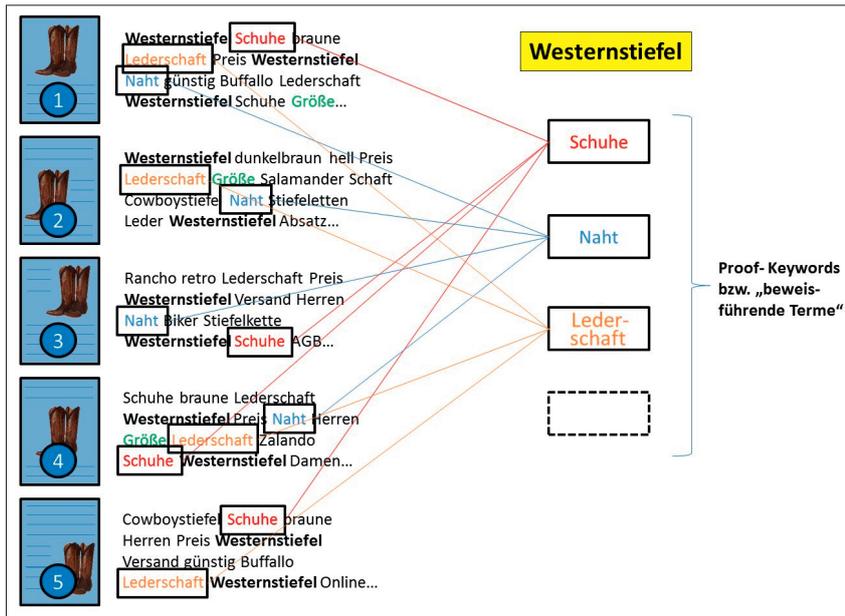


Abb. 1: Welche weiteren Worte verwenden die besten – hier fünf – Treffer? (stark vereinfachte Darstellung)

„Gegen den Mainstream kann nicht gleichzeitig auch mit dem Mainstream sein“

ist verständlich, wird aber in der Regel leider ebenso mit Nichtbeachtung ge-straft wie Webdesign, das alles möglichst anders darstellt als alle anderen.

Zum Punkt

Wie prüft man nun nach, ob ein Do-kument diesen WDF*IDF-Anforderungen genügt bzw. die richtigen Signale an Google sendet? Man braucht ein Tool. Diese sind leider noch nicht so breit gestreut und als Freeware werden sie wohl wegen des hohen Aufwands der Datenermittlung für solche Analysen auch in absehbarer Zeit nicht zur Verfügung stehen – wenn man auf Verlässlichkeit im Sinne von Datenqualität Wert legt. Am Beispiel des SEO-Tools OnPage.org kann man sich die Mechanik einer Textüberprüfung gut vor Augen führen. Dort gibt man im entsprechenden Menü ein Keyword oder eine Webadresse ein. Das Tool zieht daraufhin entweder die besten zehn Treffer bei Google zu diesem Keyword und analysiert die WDF*IDF-Werte oder eben die der eingetragenen URL (Abbildung 2).

Mittels zweier Schieberegler kann man die Anzahl der Balken vermindern und optisch vergrößern. Ein Mouseover über einem Balken zeigt die Einzelwerte an. In der Abbildung 2 erkennt man gut, dass das Wort „Westernstiefel“ bei den zehn bei Google am besten gelisteten Seiten den höchsten WDF*IDF-Wert hat („WDF*IDFMax“ ist 14,21, siehe Spalte 4 in Abbildung 2). Es kommt in allen zehn Treffern vor (Wert „10“ bei NTotal in Spalte 1 der Abb.) und dort insgesamt 245-mal (Spalte 2, Wert TFTotal). Im Maximum (WDFMax, Spalte 3) liegt der WDF-Wert, also die logarithmierte „Keyworddichte“, bei diesen zehn Treffern bei 0,63.

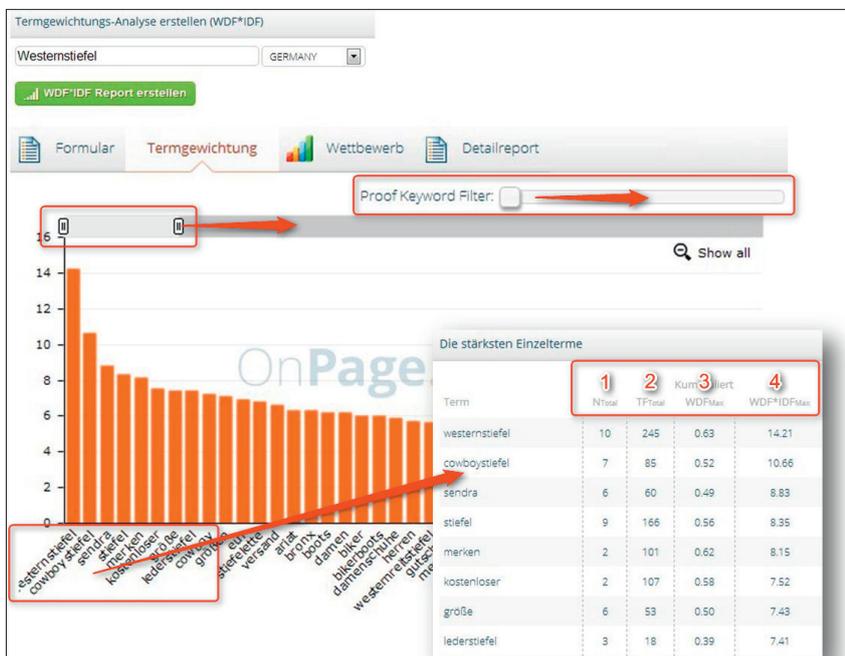


Abb. 2: Ein WDF*IDF-Report bei OnPage.org

statten, dass die technische Inhaltsanalyse, genauer die technische Erkennung des begrifflichen Schwerpunkts, möglichst treffend ist?

„Oh nein ...“, sind oft reflexartige Ausrufe bei solchen Empfehlungen, „... wir schreiben doch unsere Texte für Menschen und nicht für Google!“ Gegen diese Ehre der Schreibenden ist zwar nichts zu sagen, wird aber ein Text zum gesuchten Thema nicht gefunden, dann wird er bis auf wenige Ausnahmen auch

von niemandem gelesen und er stirbt am Ende einsam in belletristischer Schönheit. Wenn es aber umgekehrt stimmt, dass Gespräche in Vorgärten und Garten-centern neben Rasen häufig eben auch oft Worte wie grün, mähen, säen, vertikutieren, und/oder düngen enthalten, dann wird sich der Mensch möglicherweise auch beim Lesen wohlfühlen, wenn Begriffe vorkommen, die eine gewohnte und einbettende Begriffsumgebung darstellen. Textlich „anders“ sein zu wollen

Neben dem Hauptkeyword „Westernstiefel“ bringt die technische Analyse nun auch die „Beiworte“ oder auch „Proof-Keywords“ zum Vorschein. Das wären hier Begriffe wie „Cowboystiefel“, „Sendra“ (eine Markenbezeichnung), „Stiefel“, „Lederstiefel“, „Boots“, „Damen“ – aber auch „Größe“, „merken“ oder „Euro“. Nach den vorangegangenen theoretischen Überlegungen müsste man nun prüfen, ob die eigene Webseite für Westernstiefel auch möglichst viele der guten Proof-Keywords enthält und damit Suchmaschinen bei Google oder auch Bärbel aus sich selbst heraus „beweist“, dass „Westernstiefel“ nicht nur zufällig mehrmals auf der Seite vorkommt oder einzeln manuell optimiert wurde. Damit wird in gewisser Weise die textuelle Einschlägigkeit nachgewiesen.

Über den Button „Als Vorlage verwenden“ gelangt man bei OnPage.org dann in den brandneuen Textassistenten, der erst kurz vor Redaktionsschluss freigegeben wurde (Abbildung 3). Dort kann man einen Text schreiben oder einen vorhandenen einkopieren. Der Assistent gibt dann Hinweise, welche Begriffe man noch verwenden sollte (diese verschwinden, wenn sie übernommen wurden), welche man noch öfter verwenden kann und welche man bisher zu oft verwendet hat, was daher einen zu spammigen Eindruck erwecken könnte.

Sicherlich muss sich ein guter Texter nicht sklavisch an diese Vorgaben halten und auch das direkte Schreiben eines Textes in einem solchen Editor ist nicht jedermanns Sache. Sinnvoller erscheint es wohl, damit mittels Einkopieren bestehende Texte zu prüfen und gegebenenfalls entsprechend zu optimieren. In Abbildung 3 wird auch sehr gut eine sprachliche Herausforderung deutlich. Man würde in Fließtexten sicher eher von „Frauen“ als von „Damen“ sprechen. Da auf Shopdetailseiten aber natürlich klassischerweise von Schuhen für Damen (Damenschuhe ist ein feststehender Be-

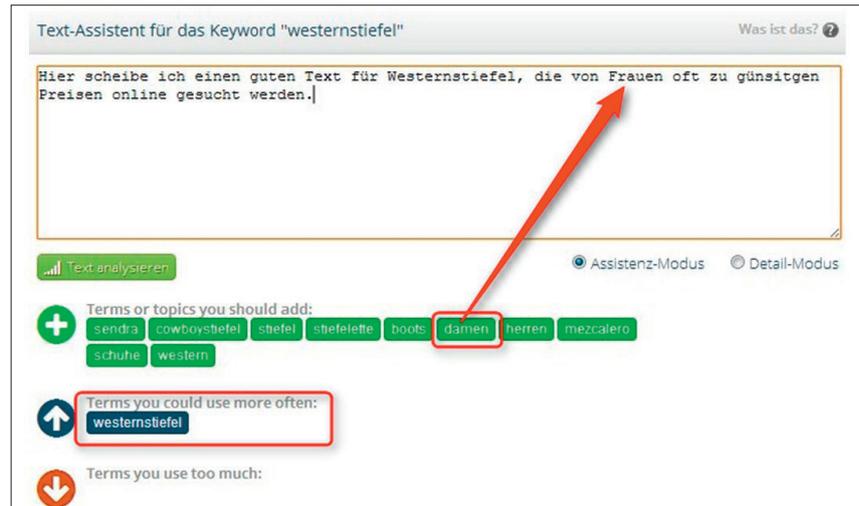


Abb. 3: Textassistenten helfen beim Texten bzw. bei Textprüfungen (Quelle: OnPage.org).

griff) und nicht von Frauenschuhen die Rede ist, mahnt das Tool an, doch das Wort „Damen“ häufiger zu verwenden. Hier ist dann Kreativität notwendig – und damit spreche ich gezielt Sie, meine verehrten Damen, als unsere Leserinnen an. Schwupps, schon ist das Wort im Text ;-)

Neben OnPage.org gibt es natürlich noch weitere Tools, die solche Analysen mehr oder weniger tief grabend anbieten. Kurz vor Redaktionsschluss hat z. B. noch Xovi die Freischaltung eines eigenen WDF*IDF-Tools angekündigt. Dort können gemäß dem Eingabeformular auch Stoppwörter und manuell zusätzlich zu analysierende URLs angegeben werden, was leider bis zum Redaktionsschluss noch nicht fehlerfrei funktioniert hat. Daher waren Aussagen dazu noch nicht möglich. Seolyze.com bietet ebenfalls entsprechende WDF*IDF-Analysen und einen Editor zum Texten an.

Hinsichtlich der Benutzerfreundlichkeit muss man sicherlich bei solchen noch relativ neuen und schnell gestarteten Tools insgesamt noch geduldig einige Abstriche machen, vor allem bei unnötigen (weil oft genügend Platz vorhanden wäre) Abkürzungen wie z. B. „OCC“, die das Interpretieren der Zahlen nicht unbedingt leichter machen und für ein einfaches Verstehen eher hinderlich sind. Bei den Toolanbietern SEOlytics und Sistrix ist ein WDF*IDF-Tool nach eigener Aussage bisher nicht geplant. Das Tool „Karlscore“ arbeitet wettbewerbsfrei und

nimmt für jede Branche daher nur einen einzigen Nutzer auf. Nach den Angaben des Anbieters Karl Kratz sind aktuell alle Slots belegt und es bestehen lange Wartelisten. Wegen dieser Nicht-Verfügbarkeit für alle anderen haben wir daher auf eine weitergehende Beschreibung des von Nutzern durchaus hoch gelobten Tools verzichtet.

Searchmetrics arbeitet derzeit nach eigener Auskunft an einem offenbar deutlich umfassenderem und weitergehendem Ansatz zur textlichen Erfassung für Optimierungszwecke, welche den eingesetzten Analyse- und Suchmethoden sowie den Crawlingmöglichkeiten von Google technologisch gesehen deutlich näher kommen soll. Das Tool könnte bereits im Sommer diesen Jahres zur Verfügung stehen. Teile der notwendigen Suchtechnologie sind ja schon als echte interne Suche auf einigen größeren Portalen im Einsatz und daher darf man gespannt sein, was Searchmetrics anders machen wird und was dies den Nutzern bringt. Entsprechende öffentliche und weiterführende Ankündigungen hat Searchmetrics in etwa für den Zeitraum angekündigt, wo diese Ausgabe der Website Boosting erscheint.

Bei der Interpretation und vor allem bei der Validität der Daten muss man derzeit nach unserer Einschätzung noch etwas vorsichtig sein. Unterschiedliche Tools werfen bei gleicher URL oder gleichen Keywords unterschiedliche Daten

aus. Dabei ist nicht auf den ersten Blick ersichtlich, ob hier mit anderen Formeln gerechnet wurde und warum oder ob sie schlicht nicht alle den nötigen Datentiefgang besitzen. Ein zuverlässiger IDF-Wert kann tatsächlich nur mit sehr aufwendigem Crawling ermittelt werden. Ebenso müssten statische Elemente wie Navigationsblöcke, Sitebars oder Footer, die auf allen Einzelseiten gleich sind, bei der Berechnung und Bewertung sicherlich anders behandelt werden, als der eigentliche Contentbereich - Google macht diese Unterschiede ja bekanntlich. Wem das alles noch zu suspekt und intransparent ist, der kann aber zumindest jetzt schon großen Nutzen aus der automatischen Ermittlung der sog. Proof-Keywords ziehen und versuchen, seine Texte mit diesen in einen vernünftigen Zusammenhang zu bringen. Selbstverständlich ist auch die Prüfung nützlich, ob denn der Hauptterm einer Seite nicht nur den höchsten WDF*IDF-Wert hat, sondern sich auch genügend stark von anderen Termen dort abhebt. Wer sein Augenmerk dabei nicht auf die absoluten Werte oder gar auf die Stellen hinter dem Komma legt, sondern die Ergebnisse mit einer gewissen Toleranz interpretiert und entsprechende Änderungen durchführt, ist sicher gut beraten.

Nach dem Punkt: Zum kritischen Fragezeichen

Ob eine gezielte und alleinige WDF*IDF-Optimierung tatsächlich so große Sprünge in den Suchergebnissen bringt, wie man es dem aktuellen Hype in der Branche entsprechend erwarten würde, konnten wir bisher nicht nachvollziehen. Unsere Hintergrundrecherchen und -gespräche brachten noch keine belastbaren Fälle ans Licht, die einer harten Nachprüfung standgehalten hätten oder die sich mit einer gewissen Eindeutigkeit darauf hätten zurückführen lassen. Zumeist hatten sich die Sitebetreiber im Zuge des neuen Mantras WDF*IDF erstmals überhaupt mit Text bzw. der Textoptimierung

„WDF*IDF ist bei SEO-Tools noch relativ neu – entsprechend wenig valide Erfahrung liegt bisher vor.“

beschäftigt oder gleichzeitig eben auch andere Optimierungsmaßnahmen vorgenommen. Einige ernst zu nehmende SEO-Experten schütteln offen oder hinter vorgehaltener Hand den Kopf und vermelden „Business as usual“. Texte hätte man schon immer nach gewissen Kriterien gestaltet und daher sei das einzig Neue, dass man das jetzt per Tool einfacher rechnen und den „Semantic Space“ (die Proof-Keywords) schneller und zuverlässiger ermitteln könne. Sieht man von diversen Beispielen ab, deren Hauptkeyword gleichzeitig auch der Domainname ist, stehen wirkliche Nachweise, dass man nur durch WDF*IDF-Optimierungen und ohne externe Links Toprankings bei Google erzielen könne, noch aus – was natürlich in keinsten Weise bedeutet, dass dies nicht funktionieren kann! Dem Vernehmen nach wundert sich derzeit wohl vor allem auch die internationale Szene über die „Crazy Germans“, die aus deren Sicht glauben, mit einfachen SEO-Formeln nach- oder vorrechnen zu können, wie man bei Google & Co ranken kann.

Bei einigen größeren Agenturen dreht man offenbar mittlerweile „am Rad“, weil viele Kunden nun alle Texte auf den WDF*IDF-Prüfziffernstand stellen wollen - bei anderen Agenturen, weil sie ihren Kunden dies gerne als weiteren, jetzt messbaren Betreuungsauftrag verkaufen möchten. Wenn man hinter etwas, was vorher nur eher gefühlsmäßig zu (be-)greifen war, plötzlich eine Zahl, noch dazu mit Kommastellen, schreiben kann, ist man im Ingenieursland Germany ja schnell und ganz besonders erregt. Die unterschiedlichen Sichtbar-

keitskennzahlen lassen grüßen.

Wäre der WDF*IDF-Wert tatsächlich ein wichtiges Rankingkriterium, dann müssten zumindest theoretisch die ersten zehn Treffer bei Google für den Suchterm jeweils den höchsten diesbezüglichen Wert ausweisen. Dies ist aber nun auch nicht immer der Fall. Und wenn dem so ist, dass auch Seiten aufgrund anderer Faktoren unter den Top Ten zu finden sind – dann beeinflussen diese doch auch die WDF*IDF-Berechnung sowie die ermittelten Proof-Keywords? Stünden beispielsweise acht von zehn Treffern wegen massiver externer Verlinkung dort, nähme man bei einer solchen textuellen Analyse dann im Extremfall nicht die völlig falschen (Text-)Signale zur Optimierung der eigenen Seite her?

Allein diese Überlegungen zeigen, dass noch nicht alle Fragen zuverlässig geklärt wurden und es weiterer Diskussionen und fachlichen Austauschs bedarf. Die [OnPage*](#)-Analyse ist und bleibt sicher und unzweifelhaft nach wie vor wichtig und birgt zum Teil hohe und oft versteckte Potenziale. Jedes Tool, das in diesem Bereich zur Optimierung hilft und Denkanstöße gibt, muss einem Sitebetreiber daher willkommen sein. Mittels der Konzentration auf einen Wert und dessen nicht ganz einfache -wirklich korrekte- Berechnung sollte man zwar keine Wunderdinge erwarten. Ihn aber umgekehrt schlicht zu ignorieren oder immer nur abzuwarten, scheint trotzdem nicht angeraten. Sich vernünftig mit Text auseinanderzusetzen – und zwar mehr als bisher –, kann nicht verkehrt sein. Man kann und darf wohl getrost davon ausgehen, dass die Toolanbieter die bisherigen ersten, durchaus zum Teil vergleichsweise beeindruckenden Würfe noch weiter professionalisieren werden, um die Aussagekraft und auch die Benutzerfreundlichkeit noch weiter zu erhöhen. Es bleibt spannend und wir bleiben daher für Sie am Ball. ¶

*siehe Online-Glossar unter www.websiteboosting.com